Régression Logistique Polytomique

Variable dépendante à K (K > 2) modalités

Ricco RAKOTOMALALA

PLAN

Les fichiers XLS associés à ce support sont disponibles en ligne

http://eric.univ-lyon2.fr/~ricco/cours/slides/multinomial_logistic_regression.xls

http://eric.univ-lyon2.fr/~ricco/cours/slides/ordinal_logistic_regression.xls

Régression logistique multinomiale

La variable dépendante est nominale à K modalités

c.-à-d. Les modalités ne sont pas ordonnées

ou On ne veut pas tenir compte d'un éventuel ordonnancement

LOGITS par rapport à une modalité de référence

Objectif : Modéliser la probabilité d'appartenance d'un individu à une catégorie « k »

$$\pi_k(\omega) = P[Y(\omega) = y_k \,/\, X(\omega)]$$

Avec toujours la contrainte

$$\sum_k \pi_k(\omega) = 1$$

Vraisemblance : On s'appuie sur le modèle multinomial

$$L = \prod_\omega [\pi_1(\omega)]^{y_1(\omega)} \times \cdots \times [\pi_K(\omega)]^{y_K(\omega)}$$

C'est bien une généralisation du modèle binomial

avec

$$y_k(\omega) = 1 \text{ssi } Y(\omega) = y_k$$

Stratégie de modélisation

On va **modéliser (K-1) rapports de probabilités** c.-à-d**. Prendre une modalité comme référence** (ex. la dernière), **et exprimer (K-1) LOGIT par rapport à cette référence** (ex. les « non-malades » à opposer à différentes catégories de maladies)

La dernière probabilité, appartenance à la K-ème catégorie, est déduite des autres

$$\pi_K(\omega) = 1 - \sum_{k=1}^{K-1} \pi_k(\omega)$$

Ré-écriture du LOGIT

Et obtention des probabilités d'appartenance

On écrit (K-1) équation LOGIT :

$$LOGIT_k(\omega) = \ln\left[\frac{\pi_k(\omega)}{\pi_K(\omega)}\right] = a_{0,k} + a_{1,k}X_1(\omega) + \cdots + a_{J,k}X_J(\omega)$$

On en déduit les (K-1) probabilités d'affectation

$$\pi_k(\omega) = \frac{e^{LOGIT_k(\omega)}}{1 + \sum_{k=1}^{K-1} e^{LOGIT_k(\omega)}}$$

Et la dernière →

$$\pi_K(\omega) = 1 - \sum_{k=1}^{K-1} \pi_k(\omega)$$

On peut vérifier que

$$\sum_{k=1}^{K} \pi_k(\omega) = 1$$

La règle d'affectation est le très classique

$$Y(\omega) = y_{k*} \Leftrightarrow y_{k*} = \arg\max_k \left[\pi_k(\omega)\right]$$

Estimation des paramètres
Estimateur du maximum de vraisemblance

La log-vraisemblance à maximiser

$$LL = \sum_{\omega} y_1(\omega) \ln \pi_1(\omega) + \cdots + y_K(\omega) \ln \pi_K(\omega)$$

Il y a (K-1) x (J+1) paramètres à estimer. On peut s'appuyer de nouveau sur la méthode de Newton-Raphson.

Vecteur gradient

(K-1)*(J+1) x 1

$$G = \begin{pmatrix} G_1 \\ \vdots \\ G_{K-1} \end{pmatrix}$$

G_k est de dim. (J+1 x 1), avec pour chaque case

$$\left[g_{k,j}\right] = \sum_{\omega} x_j(\omega) \times \left[y_k(\omega) - \pi_k(\omega)\right]$$

Matrice hessienne

(K-1)*(J+1) x (K-1)*(J+1)

$$H = \begin{pmatrix} H_{11} & & H_{1,K-1} \\ & \ddots & \\ & & H_{K-1,K-1} \end{pmatrix}$$

$H_{i,j}$ est de dim. (J+1 x J+1), définie par

$$H_{ij} = \sum_{\omega} \pi_i(\omega) \times \left[\delta_{ij}(\omega) - \pi_j(\omega)\right] \times x(\omega) x'(\omega)$$

avec $x(\omega) = \left(1, X_1(\omega), \ldots, X_J(\omega)\right)$

et $\delta_{ij}(\omega) = \begin{cases} 1, \text{si i} = \text{j} \\ 0, \text{si i} \neq \text{j} \end{cases}$

Un exemple

Choix Marque = f (sexe : femme, age) – 735 observations

Marque = {Petit prix [1], Enseigne [2], Référence du marché [3]}

Petit prix vs. Référence

$$LOGIT_1 = 22.72 - 0.47 \times F - 0.69 \times age$$

Les femmes ont moins tendance à faire confiance à la marque petit prix (par rapport à la référence)

Plus l'âge augmente, moins la personne fait le choix de la marque petit prix (par rapport à la référence)

Enseigne vs. Référence

$$LOGIT_2 = 10.95 + 0.06 \times F - 0.32 \times age$$

Classement (F = 1, age = 35)

$$LOGIT_1 = 22.72 - 0.47 \times 1 - 0.69 \times 35 \approx -1.75$$

$$LOGIT_2 = 10.95 - 0.06 \times 1 - 0.32 \times 35 \approx -0.11$$

Probabilités d'appartenance

$$\left.\begin{aligned} \pi_1 &= \frac{e^{-1.75}}{1 + [e^{-1.75} + e^{-0.11}]} = 0.08 \\ \pi_2 &= \frac{e^{-0.11}}{1 + [e^{-1.75} + e^{-0.11}]} = 0.43 \end{aligned}\right\} \quad \pi_3 = 1 - (0.08 + 0.43) = 0.49$$

La personne fera plus volontiers le choix de la marque 3 c.-à-d. la marque référence du marché

Calculer les LOGIT

Autres que par rapport à la référence

Comparer un groupe à un autre

c.-à-d. calculer

$$LOGIT_{i,j} = \ln \frac{\pi_i}{\pi_j} = \ln \frac{\pi_i / \pi_K}{\pi_j / \pi_K}$$

$$= \ln \frac{\pi_i}{\pi_K} - \ln \frac{\pi_j}{\pi_K} = LOGIT_i - LOGIT_j$$

Confronter « Petit prix [1] » et « Enseigne [2] »

Petit prix vs. Référence $\boxed{LOGIT_1 = 22.72 - 0.47 \times F - 0.69 \times age}$

Enseigne vs. Référence $\boxed{LOGIT_2 = 10.95 + 0.06 \times F - 0.32 \times age}$

Petit prix vs. Enseigne $\boxed{LOGIT_{1,2} = 11.77 - 0.52 \times F - 0.37 \times age}$

Les femmes ont moins tendance à faire confiance à la marque petit prix (par rapport à « enseigne »)

Plus l'âge augmente, moins la personne fait le choix de la marque petit prix (par rapport à « enseigne »)

Évaluation globale de la régression

Au delà des pseudo-R² et Matrice de confusion - Test du rapport de vraisemblance

Principe, toujours le même, comparer le modèle avec le modèle trivial composé de la seule constante (dans chaque équation LOGIT)

$$\hat{a}_{0,k} = \ln \frac{n_k}{n_K}$$

Statistique du test

$$LR = -2\ln\frac{L(0)}{L(M)} = -2 \times LL(0) - \left[-2LL(M)\right]$$

Distribution

$$LR \equiv \chi^2\left[\left(n-(1)\times(K-1)\right)-\left(n-(J+1)\times(K-1)\right)\right] \equiv \chi^2\left[J\times(K-1)\right]$$

Predicted attribute		brand
Ref. value		M_Reference
Number of examples		735
Model Fit Statistics		
Criterion	Intercept	Model
AIC	1595.792	1417.941
SC	1604.991	1445.541
-2LL	1591.792	1405.941
Model Chi² test (LR)		
Chi-2		185.8502
d.f.		4
P(> Chi-2)		0.0000
R²-like		
McFadden's R²		0.1168
Cox and Snell's R²		0.2234
Nagelkerke's R²		0.2524

$$LR = 1591.762 - 1405.941 = 185.85$$
$$p - value < 0.0001$$

Les variables prédictives emmènent significativement de l'information dans l'explication des valeurs prises par la variable dépendante.

Pseudo-R² :
= 0 → modèle inutile
=1 → modèle parfait

Évaluation individuelle des coefficients (dans chaque LOGIT)

Test de Wald

Tester dans une équation LOGIT : le coefficient est significativement différent de 0 ?

$$\begin{cases} H_0 : a_{j,k} = 0 \\ H1 : a_{j,k} \neq 0 \end{cases}$$

Statistique du test de Wald

Elle suit une loi du KHI-2 à 1 d.d.l

$$W_{j,k} = \left(\frac{\hat{a}_{j,k}}{\hat{\sigma}_{\hat{a}_{j,k}}} \right)^2$$

La matrice de Variance-covariance, inverse de la matrice Hessienne, fournit les estimations des écarts type.

Class.Value	M_PetitPrix.1				M_Enseigne.2			
Pred.Att.	Coef.	Std.Err	Wald	p-value	Coef.	Std.Err	Wald	p-value
intercept	22.721397	-	-	-	10.946741	-	-	-
femme	-0.465941	0.2261	4.247	0.0393	0.057873	0.1964	0.08681	0.7683
age	-0.685908	0.0626	120	0.0000	-0.317702	0.0440	52.12	0.0000

Sortie du logiciel

Calcul de la matrice Hessienne

Hessian	Intercept.1	Femme.1	Age.1	Intercept.2	Femme.2	Age.2
Intercept.1	124.55	75.15	3994.78	-84.52	-51.68	-2688.56
Femme.1	75.15	75.15	2405.87	-51.68	-51.68	-1641.70
Age.1	3994.78	2405.87	128479.71	-2688.56	-1641.70	-85734.53
Intercept.2	-84.52	-51.68	-2688.56	173.70	112.66	5716.17
Femme.2	-51.68	-51.68	-1641.70	112.66	112.66	3704.70
Age.2	-2688.56	-1641.70	-85734.53	5716.17	3704.70	188915.35

$$\hat{\Sigma} = H^{-1}$$

INV--Var.Covar	Intercept.1	Femme.1	Age.1	Intercept.2	Femme.2	Age.2
Intercept.1	4.2355	-0.0572	-0.1284	1.6579	-0.0347	-0.0480
Femme.1	-0.0572	0.0511	0.0008	-0.0265	0.0260	0.0003
Age.1	-0.1284	0.0008	0.0039	-0.0485	0.0005	0.0014
Intercept.2	1.6579	-0.0265	-0.0485	2.2295	-0.0396	-0.0653
Femme.2	-0.0347	0.0260	0.0005	-0.0396	0.0386	0.0004
Age.2	-0.0480	0.0003	0.0014	-0.0653	0.0004	0.0019

Racine(Variance)						
	2.0580					
		0.2261				
			0.0626			
				1.4932		
					0.1964	
						0.0440

Détails calcul Wald

$$\hat{\sigma}^2_{\hat{a}_{femme,1}} = 0.0511$$

$$\hat{\sigma}_{\hat{a}_{femme,1}} = 0.2261$$

$$W_{femme,1} = \left(\frac{-0.466}{0.2261} \right)^2 = 4.247$$

$$p-value_{femme,1} = 0.0393$$

Évaluation globale des coefficients (dans l'ensemble des LOGIT)

Généralisation du test de Wald sur plusieurs coefficients

Tester : la variable joue un rôle dans le problème traité ?

$$\begin{cases} H_0 : \forall k, a_{j,k} = 0 \\ H1 : \exists k, a_{j,k} \neq 0 \end{cases}$$

Le coefficient de la variable j est-il égal à zéro dans toutes les équations ?

Statistique du test de Wald, elle suit une loi du KHI-2 à (K-1) d.d.l

$$W_j = \hat{a}'_j \, \hat{\Sigma}_j^{-1} \hat{a}_j \cong \chi^2(K-1)$$

avec

$$\hat{a}_j = \begin{pmatrix} \hat{a}_{j,1} \\ \hat{a}_{j,2} \\ \vdots \\ \hat{a}_{j,K-1} \end{pmatrix}_{(K-1)\times 1}$$

et

$$\hat{\Sigma}_j$$

$(K-1)\times(K-1)$

Est la partie de la matrice de variance co-variance relative au coefficient de la variable Xj

Rappel – Mat. Var.Covar.

INV--Var.Covar	Intercept.1	Femme.1	Age.1	Intercept.2	Femme.2	Age.2
Intercept.1	4.2355	-0.0572	-0.1284	1.6579	-0.0347	-0.0480
Femme.1	-0.0572	0.0511	0.0008	-0.0265	0.0260	-0.0003
Age.1	-0.1284	0.0008	0.0039	-0.0485	0.0005	0.0014
Intercept.2	1.6579	-0.0265	-0.0485	2.2295	-0.0396	-0.0653
Femme.2	-0.0347	0.0260	0.0005	-0.0396	0.0386	0.0004
Age.2	-0.0480	-0.0003	0.0014	-0.0653	0.0004	0.0019

Rappel – Coef.(femme) LOGIT 1 et 2

$$\hat{a}_{femme} = \begin{pmatrix} -0.466 \\ +0.058 \end{pmatrix}$$

$$\hat{\Sigma}_{femme} = \begin{pmatrix} 0.0511 & 0.0260 \\ 0.0260 & 0.0386 \end{pmatrix} \rightarrow \hat{\Sigma}^{-1}_{femme} = \begin{pmatrix} 29.75 & -20.03 \\ -20.03 & 39.41 \end{pmatrix} \rightarrow W_{femme} = 7.6704$$

Avec un $\chi^2(2)$ → p-value = 0.0216

À 5%, la variable est globalement significative, on doit la conserver.

Interprétation des coefficients
Odds-ratio – Marque = f(femme)

Croisement Marque vs. Femme

Nombre de brand	femme		
brand	1	0	Total
M__PetitPrix	115	92	207
M_Enseigne	208	99	307
M_Reference	143	78	221
Total	466	269	735

Odds(Référence)

M__PetitPrix	0.804196	1.179487
M_Enseigne	1.454545	1.269231

Odds-Ratio (Référence)

M__PetitPrix	0.681818
M_Enseigne	1.146006

Résultats de la régression logistique

Class.Value	M__PetitPrix				M_Enseigne		
Pred.Att.	Coef.	Std.Err	Wald	p-value	Coef.	Std.Err	Wald
constant	0.16508	-	-	-	0.238411	-	-
femme	-0.38299	0.1984	3.725	0.0536	0.136282	0.1863	0.5349

$$OR_j = e^{\hat{a}_j}$$

Odds-Ratio	Femme
M_PetitPrix	**0.68182**
M_Enseigne	**1.14601**

Pour une variable nominale (0/1)

- L'interprétation en termes de surcroît de risque (odds-ratio) reste valable
- Mais par rapport à la modalité de référence (ex. une femme a 0.6818 fois plus de chances qu'un homme de préférer la marque « petit prix » à la marque « référence »… hum !!! La lecture est difficile à ce stade…).

Pour les autres types de variables (nominale à + de 2 catégories, ordinale, continue), les interprétations vues pour la rég.logistique binaire restent valables, mais toujours par rapport à la modalité de référence.

Remarque : L'exponentielle des constantes des régression de la régression s'interprètent comme des ODDS (ex. Pour la modalité « Petit prix » → EXP(0.16508) = 1.179487 c.-à-d. *chez les hommes [femme = 0], on a 1.17 fois plus de chances de choisir la marque petit prix que la marque de référence*).

Régression logistique ordinale (1)

La variable dépendante est ordinale à K modalités

Plus que la prédiction, c'est l'interprétation des coefficients qui importe

LOGITS adjacents

Définition des LOGITS adjacents

Principe :

Calculer le LOGIT du passage d'une catégorie à l'autre

(K-1) équations LOGIT sont calculés → (K-1) x (J+1) paramètres à estimer

$$\begin{cases} LOGIT_1(\omega) = \ln\left[\dfrac{\pi_1(\omega)}{\pi_2(\omega)}\right] = a_{0,1} + a_{1,1}X_1(\omega) + \cdots + a_{J,1}X_J(\omega) \\ \cdots \\ LOGIT_{K-1}(\omega) = \ln\left[\dfrac{\pi_{K-1}(\omega)}{\pi_K(\omega)}\right] = a_{0,K-1} + a_{1,K-1}X_1(\omega) + \cdots + a_{J,K-1}X_J(\omega) \end{cases}$$

Même idée que le modèle multinomial, sauf que la catégorie de référence change à chaque étape.

On évalue le passage de la modalité (k) à (k-1) [moins bonne]

Cette écriture peut être vue comme une ré-interprétation du modèle multinomial !!!

$$\begin{cases} \ln\left[\dfrac{\pi_2(\omega)}{\pi_1(\omega)}\right] = -LOGIT_1(\omega) \\ \ln\left[\dfrac{\pi_3(\omega)}{\pi_1(\omega)}\right] = -LOGIT_2(\omega) - LOGIT_1(\omega) \\ \cdots \\ \ln\left[\dfrac{\pi_K(\omega)}{\pi_1(\omega)}\right] = -LOGIT_{K-1}(\omega) - \cdots - LOGIT_2(\omega) - LOGIT_1(\omega) \end{cases}$$

On peut utiliser les résultats du modèle multinomial pour estimer les paramètres.

Les évaluations et les tests de significativité sont les mêmes.

Voyons ce qu'il en est pour l'interprétation des coefficients.

Un exemple

Qualité du vin = f (soleil,pluie) – Données centrées réduites

-2LL(0)	74.647
-2LL(M)	33.635
LR	41.012
d.f.	4
p-value	0.0000

Modèle globalement significatif

Coefficients des LOGITS

Sans contraintes								
	bad => medium				**medium => good**			
	Coef.	Std.Err	Wald	p-value	Coef.	Std.Err	Wald	p-value
Intercept	1.655	-	-	-	-1.706	-	-	-
soleil	4.051	1.982	4.178	0.0409	2.478	1.128	4.827	0.0280
pluie	-1.702	1.038	2.689	0.1010	-1.561	1.024	2.327	0.1272

Comment lire ces résultats ?

bad → medium

Le nombre de jours d'ensoleillement augmente de 1 é.t. → le LOGIT augmente de 4.051

→

ODDS-RATIO = exp(4.051) = 57.5

Il y a 57 fois plus de chances que le vin soit « medium » que « bad »

medium → good

Le nombre de jours d'ensoleillement augmente de 1 é.t. → le LOGIT augmente de 2.478

→

ODDS-RATIO = exp(2.478) = 11.9

Il y a 11 fois plus de chances que le vin soit « good » que « medium »

!

- L'interprétation dépend de la valeur (du niveau) de la variable réponse (Y)
- Lecture difficile dès que le nombre de variables augmente

LOGITS adjacents à coefficients constants

Réduire l'espace des solutions pour une autre (meilleure ?) interprétation des résultats

Introduire contrainte suivante

$$LOGIT_k(\omega) = \ln\left[\frac{\pi_k(\omega)}{\pi_{k+1}(\omega)}\right] = a_{0,k} + a_1 X_1(\omega) + \cdots + a_J X_J(\omega$$

Le surcroît de risque induit par une variable explicative pour passer d'une catégorie à la suivante ne dépend pas du niveau (de la valeur) de la variable dépendante.

Modèle globalement significatif

-2LL(0)	74.647
-2LL(M)	34.252
LR	40.396
d.f.	2
p-value	0.0000

Coefficients des LOGITS

Avec contrainte				
	Coef.	Std.Err	Wald	p-value
Intercept (bad => medium)	1.287	-	-	-
Intercept (medium => good)	-2.042	-	-	-
soleil	3.035	0.981	9.568	0.0020
pluie	-1.452	0.695	4.362	0.0368

Comment lire ces résultats ?

Les constantes : Lorsque l'on observe la valeur moyenne des explicatives (soleil = 0, pluie = 0, !!! Les variables sont centrées), on a **EXP(1.287) = 3.6** fois plus de chances d'avoir un vin **medium que bad** ; et EXP(-2.042) = 0.13 plus de chances d'avoir un vin good que medium (ou 1/13 = 7.7 fois plus de chances d'avoir un vin medium que good).

Les coefficients des variables (ex. Soleil) : Lorsqu'on augmente de 1 écart type l'ensoleillement, on EXP(3.035) = 20.8 fois plus de chances de passer dans la catégorie supérieure (que l'on soit présentement bad ou medium !!!).

Plus de contraintes → Plus de stabilité : On réduit les écarts type des estimateurs, certaines variables peuvent devenir significatives (ex. pluie : significative à 5% maintenant)

! Résultat plus concis, plus lisible

Mais la vision reste locale : on évalue le risque de passer d'un niveau à l'autre

Régression logistique ordinale (2)

La variable dépendante est ordinale à K modalités

Plus que la prédiction, c'est l'interprétation des coefficients qui importe

ODDS-RATIO cumulatifs

Définition des LOGITS cumulatifs

Principe : Calculer le LOGIT d'être au delà ou en deçà du niveau y_k de la variable Y

Définissons la probabilité cumulée $P(Y \leq k / X) = \pi_1 + \cdots + \pi_k$

LOGITS cumulatifs $LOGIT_k = \ln\left(\frac{P(Y \leq k / X)}{P(Y > k / X)}\right) = \ln\left(\frac{P(Y \leq k / X)}{1 - P(Y \leq k / X)}\right) = \ln\left(\frac{\pi_1 + \cdots + \pi_k}{\pi_{k+1} + \cdots + \pi_K}\right)$

(K-1) équations LOGIT sont définis (dans un premier temps)

$$\begin{cases} LOGIT_1(\omega) = a_{0,1} + a_{1,1} X_1(\omega) + \cdots + a_{J,1} X_J(\omega) \\ \cdots \\ LOGIT_{K-1}(\omega) = a_{0,K-1} + a_{1,K-1} X_1(\omega) + \cdots + a_{J,K-1} X_J(\omega) \end{cases}$$

Introduisons de nouveau l'hypothèse : le rôle d'une variable ne dépend pas du niveau de Y

$$LOGIT_k(\omega) = a_{0,k} + a_1 X_1(\omega) + \cdots + a_J X_J(\omega)$$

- Modèle à Odds ratio cumulatifs proportionnels : **très utilisé** car d'excellente interprétation
- Lien sous-jacent entre les Xj et une variable latente Z (version continue de Y)
- Constantes = log-odds d'être inf. ou égal à la catégorie k lorsque X1=…=XJ=0 → sorte de paliers
- Les droites LOGIT vs. Explicatives sont donc parallèles, décalées selon les constantes
- (K + J – 1) paramètres à estimer

Probabilités et vraisemblance

Calculer la probabilité de se situer en deçà d'une valeur de Y (pour un individu)

$$P(Y \le k / X) = \frac{e^{a_{0,k} + a_1 X_1 + \cdots + a_J X_J}}{1 + e^{a_{0,k} + a_1 X_1 + \cdots + a_J X_J}}$$

Vision globale c.-à-d. situer le positionnement d'un individu sur l'échelle de valeur définie par Y

$$\left.\begin{array}{l} P(Y > k / X) = 1 - P(Y \le k / X) \\ P(Y = k / X) = P(Y \le k / X) - P(Y \le k-1 / X) \\ P(Y \le K / X) = 1 \end{array}\right\} \text{Bien évidemment}$$

La fonction de vraisemblance à maximiser

$$L = \left[\prod_{k=1}^{K-1} \prod_{\omega : Y(\omega) = k} P\big(Y(\omega) \le k / X(\omega)\big) \right] \left[\prod_{\omega : Y(\omega) = K} P\big(Y(\omega) = K / X(\omega)\big) \right]$$

$\dashrightarrow$ Pour obtenir les coefficients $(\hat{a}_{0,1}, \ldots, \hat{a}_{0,K-1}, \hat{a}_1, \ldots, \hat{a}_J)$

Exemple des Vins

-2LL(0)	74.647
-2LL(M)	34.441
LR	40.207
d.f.	2
p-value	0.0000

Modèle globalement significatif

Coefficients des LOGIT

LOGIT cumulatifs proportionnels				
	Coef.	Std.Err	Wald	p-value
Intercept (<= bad)	-1.427	-	-	-
Intercept (<= medium)	2.322	-	-	-
soleil	-3.226	0.943	11.716	0.0006
pluie	1.567	0.730	4.603	0.0319

Comment lire ces résultats ?

Classement d'une observation :

Soleil = 1 et pluie = 1

$$LOGIT_{bad} = -1.427 - 3.226 \times 1 + 1.567 \times 1 = -3.086$$

$$\Rightarrow P(Y \leq bad / X) = \frac{e^{-3.086}}{1 + e^{-3.086}} = 0.044$$

$$P(Y = bad / X) = 0.044$$

$$LOGIT_{medium} = 2.322 - 3.226 \times 1 + 1.567 \times 1 = 0.662$$

$$\Rightarrow P(Y \leq medium / X) = \frac{e^{0.662}}{1 + e^{0.662}} = 0.660$$

$$P(Y = medium / X) = 0.660 - 0.044 = 0.616$$

$$P(Y = good / X) = 1 - 0.616 = 0.384$$

Les constantes : Lorsque l'on observe la valeur moyenne des explicatives (soleil = 0, pluie = 0, !!! Les variables sont centrées), on a **EXP(-1.427) = 0.24** fois plus de chances d'avoir un vin **bad que meilleur que bad {medium ou good}** ; et EXP(2.322) = 10.2 fois plus de chances d'avoir un vin inf. ou égal à medium que good.

Exemple des vins (suite)

Coefficients des LOGIT

LOGIT cumulatifs proportionnels				
	Coef.	Std.Err	Wald	p-value
Intercept (<= bad)	-1.427	-	-	-
Intercept (<= medium)	2.322	-	-	-
soleil	-3.226	0.943	11.716	0.0006
pluie	1.567	0.730	4.603	0.0319

Comment lire ces résultats ?

Les coefficients (ex. pour la variable soleil) : Lorsque l'ensoleillement augmente de Δ, tous les LOGIT augmentent de (toutes choses égales par ailleurs)

$$\forall k : LOGIT_k(X_j + \Delta) - LOGIT_k(X_j) = \hat{a}_j \times \Delta$$

En termes d'odds-ratio

$$OR_j = \frac{P(Y \leq k / X_j + \Delta) / P(Y > k / X_j + \Delta)}{P(Y \leq k / X_j) / P(Y > k / X_j)} = e^{\hat{a}_j \times \Delta}$$

Lorsque les jours d'ensoleillement augmentent de 1 é.t., on a EXP(-3.226) = 0.04 fois plus de chances d'avoir un vin moins bon qu'il ne l'est (ou mieux, 1/0.04 #25 fois plus de chances d'améliorer sa qualité). **Ceci quel que soit son niveau de qualité actuel** !!!

Pourquoi « proportionnel » ? L'odds-ratio cumulatif, suite à une variation Δ d'une des variables explicatives Xj, est proportionnel à Δ et ne dépend pas du niveau de Y. Le paramètre a_j de la variable est le coefficient de proportionnalité.

Exemple 2 – Hypertension

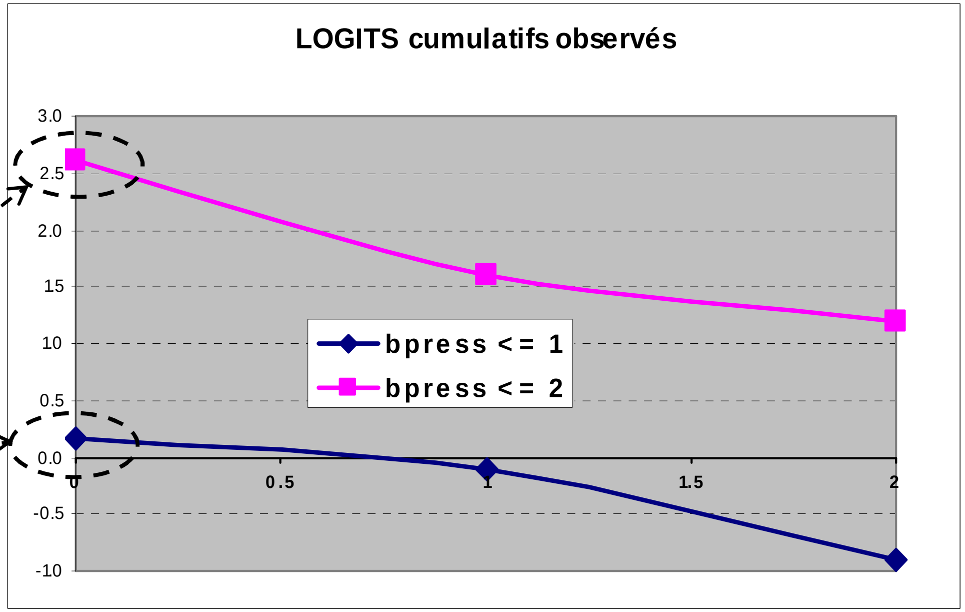
Comprendre les LOGITS cumulatifs proportionnels

Hypertension {0, 1, 2} = f [Surpoids {1, 2, 3}]

Nombre de bpress	overweight			
bpress	0	1	2	Total
3	11	13	37	61
2	63	28	76	167
1	88	37	46	171
Total	162	78	159	399

$$\ln\frac{P(Y\le 2/0)}{P(Y>2/0)}=\ln\frac{63+88}{11}=2.6194$$

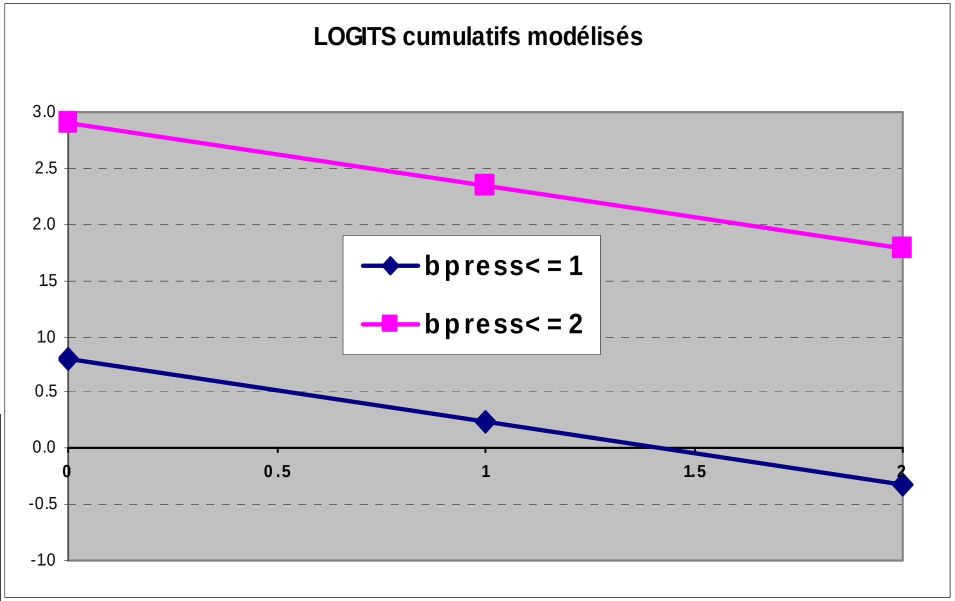
$$\ln\frac{P(Y\le 1/0)}{P(Y>1/0)}=\ln\frac{88}{63+11}=0.1733$$



Hypothèses pour la modélisation : (1) ces courbes sont des droites ; (2) elles sont parallèles.

-2LL(0)	809.806
-2LL(M)	782.286
LR	27.520
d.f.	1
p-value	0.0000

LOGIT cumulatifs proportionnels				
	Coef.	Std.Err	Wald	p-value
Intercept (<= 1)	0.805	-	-	-
Intercept (<= 2)	2.912	-	-	-
overweight	-0.561	0.109	26.328	0.0000



(1) Effet du surpoids sur la tension bien appréhendé

(2) Ajustement pas très bon (cf. Pseudo-R²)

(3) On peut tester l'hypothèse « LOGITS parallèles »

(Cf. Score test for proportional odds assumption dans SAS par ex.)

(On pourrait aussi utiliser le test de rapport de vraisemblance)

En ligne
A. SLAVKOVIC – « STAT 504 – Analysis of discrete data »
http://www.stat.psu.edu/~jglenn/stat504/08_multilog/01_multilog_intro.htm
Très riche avec des exemples sous SAS et R (redirection sur les sites web)

Ouvrages
J.P. NAKACHE et J. CONFAIS – « Statistique Explicative Appliquée », TECHNIP, 2003.